MOLECULAR AND CELLULAR BIOLOGY, Mar. 1993, p. 1583–1589
0270-7306/93/031583-07$02.00/0

Vol. 13, No. 3

Complex Telomere-Associated Repeat Units in Members of the Genus *Chironomus* Evolve from Sequences Similar to Simple Telomeric Repeats

LENA NIELSEN AND JAN-ERIK EDSTRÖM*

Department of Molecular Genetics, University of Lund, Sölvegatan 29, S-22362 Lund, Sweden

Received 29 September 1992/Returned for modification 9 November 1992/Accepted 15 December 1992

The dipteran *Chironomus tentans* has complex tandemly repeated 350-bp DNA sequences at or near the chromosome ends. As in *Drosophila melanogaster*, short simple repeats with cytosines and guanines in different strands have never been observed. We were therefore interested in learning whether the *Chironomus* repeats could have evolved from simple sequence telomeric DNA, which might suggest that they constitute a functional equivalent. We screened for repeat units with evolutionarily ancient features within the tandem arrays and recovered two clones with a less-evolved structure. Sequence analysis reveals that the present-day 350-bp unit probably evolved from a simpler 165-bp unit through the acquisition of transposed sequences. The 165-bp unit contains DNA with a highly biased distribution of cytosine and guanine between the two strands, although with the ratios inverted in two minor parts of the repeat. It is largely built up of short degenerate subrepeats for which most of the sequence can be reconstructed. The consensus for the subrepeat sequence is similar to the simple telomeric repeat sequences of several kinds of eukaryotes. We propose that the present-day unit has evolved from telomeric, simple sequence, asymmetric DNA from which it has retained some original sequence features and possibly functions.

At the ends of most eukaryotic chromosomes, there are short, simple DNA repeats, slightly different in various lower and higher eukaryotes (reviewed in reference 5). Such DNA of different origin has one feature in common: all or most of the cytosines are in one strand and the guanines in the other. The terminal DNA can be added to the ends enzymatically by telomerase (17, 18) or, as shown in yeast cells, by gene conversion (33). No such DNA has been found in dipteran species. The telomere-associated (TA) DNA recovered in *Drosophila* species (1, 3, 4, 26, 28, 34) and *Chironomus* species (8, 9, 24, 29) is more complex, like subtelomeric DNA in other species. It will be difficult to exclude that a few short repeats, of potential functional importance, remain undetected at the ends of dipteran chromosomes. An alternative, however, is that their complex repeats have taken over the function of the short repeats and are truly terminal. If so, dipteran repeats may have evolved from simple repeats. Here we pose the question whether complex repeats can be traced back in evolution to simple telomeric repeats.

In chironomids, DNA at or close to the chromosome ends, TA DNA, consists of regular tandem arrays (8, 9, 24, 29). Similar DNA has also been described for some *Drosophila* species (1). Because of the well-defined structure of the 350-bp TA repeats in the subgenus *Camptochironomus*, with two pairs of subrepeats in an alternating arrangement, separated by four linker regions, nonrepetitive within the 350-bp unit, the TA repeats provide an interesting model to learn how a complex unit is formed during evolution from a simpler organization. We have earlier asked whether the formation of subrepeats, which show more than 90% mutual homology, is due to differences in rates of evolution along the 350-bp repeat unit after previous duplication of an approximately 175-bp half-repeat. Alternatively, forces exist that give members of a pair of subrepeats a parallel evolution. We have compared repeat units in the closely related *Chironomus tentans* and *C. pallidivittatus* and concluded that mutations are preferentially located in linker regions, thus supporting the first alternative (24).

From the interspecies comparison, we learned what happened after speciation, when most of the morphology of the repeat unit is already established. For present purposes, we searched for a repeat unit with early evolutionary characters. For tandem repeats, this is not necessarily unrealistic, since tandem repeats may contain units at their ends representing earlier evolutionary phases (12, 14, 19). We argued that we might be able to select for such variants by searching for repeats not containing a specific cluster of mutations, known from the interspecies study to have been formed late during evolution. This strategy was successful and led to the recovery of two clones, which both had features expected for precursor forms, i.e., with less difference between the two 175-bp halves. In this case, two of the linkers that are normally highly different within the 350-bp unit were almost identical. Closer study of the structure of the DNA before and after the differentiation of the linkers allowed us to deduce what processes might have led to the differentiated 350-bp unit. In this way, we obtained the structure of a putative primordial 165-bp unit. This unit contains DNA that is highly asymmetric in distribution of G and C between the two strands. However, two minor regions of the unit have an inverted distribution of G and C between the strands compared with the remainder. We propose that the complex TA repeat units are formed from simple sequence DNA with the Cs in one strand and the Gs in the other mainly by inversions and transpositions. The simpler 165-bp repeat largely consists of 7- to 10-mer degenerate subrepeats having a consensus sequence with striking similarity to telomeric short repeats of different eukaryotes. The asymmetric feature of this DNA remains in part of the present-day 350-bp unit. The

* Corresponding author.

differentiation of the unit presumably has occurred to permit acquisition of new functions, as will be discussed.

MATERIALS AND METHODS

Genomic DNA from *C. tentans* was prepared as described previously (29) and digested with *Nla*III. A digest of 3 μg of DNA was separated in a 1% agarose gel, and the 350-bp region was extracted from the gel. The eluted fragments were cloned into the *Sph*I site of pUC18. The library was screened with the *C. tentans* standard repeat (24), representing the whole repeat unit, and with the oligonucleotide 5′-CAATTT CTCAACCAATTTGGGATGTTTATATATCGA-3′, specific for a short linker region existing in *C. tentans* but not in *C. pallidivittatus*. Colony hybridizations were done on nylon membranes (Hybond-N) by standard protocols (21). Hybridization with DNA probe was in 5× SSPE (1× SSPE is 0.18 M NaCl, 10 mM $NaPO_4$, and 1 mM EDTA [pH 7.7])–0.5% (wt/vol) sodium dodecyl sulfate (SDS)–5× Denhardt's solution–100 μg of salmon sperm DNA per ml at 65°C. Washes were made in 1× SSC (0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS at 65°C. Hybridization conditions for the oligonucleotide were as above, but at 55°C, and washes were done in 2 ×SSC–0.1% SDS at 55°C. The inserted DNA was sequenced by the dideoxy chain termination method (30).

Hybridization with the degenerate 15-mer oligonucleotide was done as for the longer oligonucleotide but at room temperature like the washings.

Nucleotide sequence accession numbers. The GenBank accession number for the 342-bp clone A is L04934 and for the 368-bp clone B is L04935.

RESULTS

Structure of TA repeat unit. In *C. pallidivittatus*, the repeat unit consists of two pairs of subrepeats, Ia-Ib and IIa-IIb, in an alternating arrangement. These are separated by linker regions, L1 to L4, which are nonrepetitive within the unit (29). The order along the repeat unit is Ia-L1-IIa-L2-Ib-L3-IIb-L4 (Fig. 1b). In the sibling species *C. tentans* (Fig. 1a), a cluster of mutations within IIb subdivides this unit into two parts with an intervening linker region (24). By analogy, IIa also becomes similarly subdivided since the original homolog is no longer repeated within the 350-bp unit.

Screen for nonstandard repeat units. *Nla*III digests almost all the genomic *C. tentans* TA DNA to monomeric units, with a size of 350 bp (24). Such fragments were used for a library, and 47 positive clones were obtained in a screen with the whole repeat unit. Further screening was done with the oligonucleotide for the linker region specific for *C. tentans*, subdividing the subrepeat IIb. This screen yielded 41 positive clones, and the remaining 6 negative clones were used in a search for repeats with ancient features. Three clones had given weak signals in the first screen and were shown by sequencing not to have any traits typical for TA DNA. Three clones had, however, features of a less-evolved TA DNA. Here, two regions that are linkers in the standard type of repeat were almost identical. Consequently, the 350-bp repeat unit in these three clones consists of one pair of long subrepeats separated on both sides by linker regions (Fig. 1c and d). Two of the three clones were identical. One of them represented DNA situated at the flank of a repeat array and was joined to non-TA DNA. We denote the clone lacking flanking sequences as clone A (Fig. 1c) and the clone containing flanking DNA as clone B (Fig. 1d). The term

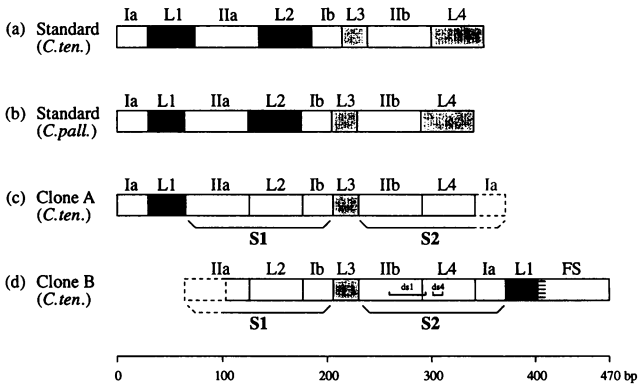

FIG. 1. Diagrams of different TA repeats from *Chironomus* species. White boxes show subrepeats, and grey boxes show linker regions. Light grey fields within IIa and IIb for C. tentans (*C. ten.*) in panel a indicate short linker regions specific for this species. In clone A, the cloning site is within L3, and the clone is shown in a circular permutation in which it starts with subrepeat Ia. The dashed box downstream of L4 representing Ia in panel c reiterates the Ia subrepeat at the start of the diagram in order to clarify the arrangement of the large S1 and S2 subrepeat units. Clone B is shown in panel d as it was cloned, and the dashed box at the start indicates the part of IIa that was not included in the clone. The unsharp border between L1 and flanking sequence (FS) is indicated by dashes. The positions of the ds1 and ds4 regions in clone B are shown. Standard, standard repeat; *C. pall.*, *C. pallidivittatus*.

standard repeat is used for the sequence characteristic for the main part of the TA DNA in *C. tentans* (24).

Sequence of clone A. To facilitate comparisons with other units, clone A is presented starting with subrepeat Ia, although the *Nla*III cloning site is in L3 (Fig. 2). Compared with the standard repeat, the main difference is in the linker region L2, which here is almost identical to L4. Both L2 and L4 are consequently not yet linkers but parts of large subrepeats: L2 is part of a subrepeat encompassing IIa-L2-Ib, here designated S1, and L4 is part of an S2 subrepeat with the structure IIb-L4-Ia (Fig. 1c and d). The order along the repeat consequently is S1-L3-S2-L1. During evolution, a putative transposition has changed the structure of L2 in its distal 39-bp extension. In clone A, a possible target site, ATTCGT (underlined in Fig. 2), occupies positions 8 to 13 in L2. In the standard repeat, the site is duplicated, and the 5′-proximal site is in the same location and the 5′-distal site is at the end of L2 (both underlined). In parallel with the assumed transposition, a deletion of a 1-bp-longer fragment of the original L2 has occurred. The inserted DNA is rich in A at its 3′ end, suggesting transposition of a reversely transcribed mRNA fragment.

(As a result of the changes in L2, most of it shows little identity with L4 in the standard repeat. There is, nevertheless, a short sequence in L2 of 6 bp immediately following the first duplicated target site where a cluster of mutations has changed L4 in positions 313 to 318 from CTGCAA to AGCTAT, making it similar to L2. This suggests a gene conversion event that extended the homology in the region upstream of the putative transposition in the 3′ direction.)

Also in other respects, clone A shows less differentiation between the two halves of the repeat unit. Thus, there is only a 1-bp difference between Ia and IIa on one side and Ib and IIb on the other, whereas in the standard repeat, there are six differences. This supports the view that clone A shows traits of evolutionarily early forms.

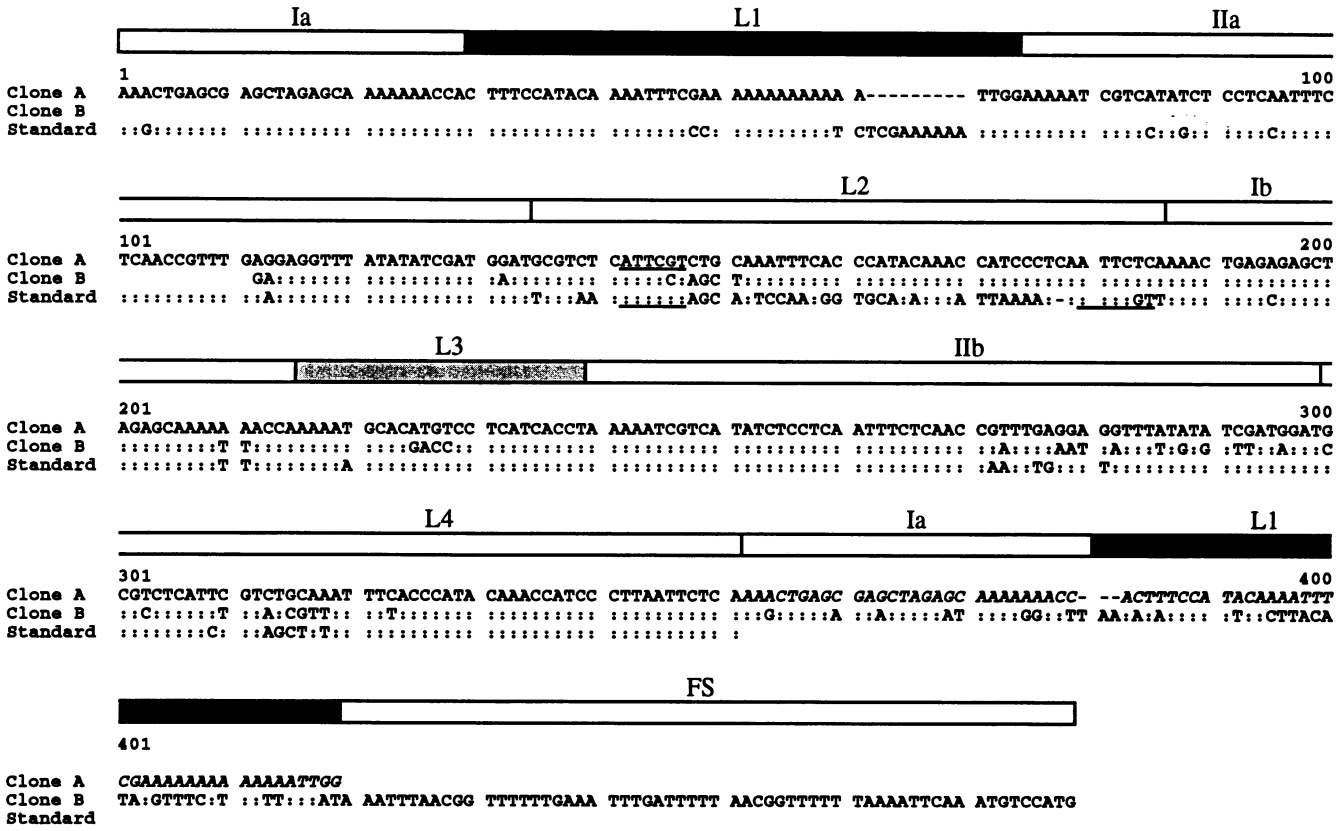

FIG. 2. Sequences in the 5′-3′ direction showing the C strand of clone A (top line) in a circular permutation as indicated in the legend to Fig. 1. Subrepeat Ia and linker region L1 have been reiterated at the end of the sequence (in italics) to permit comparison with clone B, which is shown under the standard repeat from *C. tentans* on the middle line. Identities with the top sequence (clone A) are indicated with colons.

A final difference, used in the screen for the repeat variants, is the absence in clone A of the cluster of mutations giving a short degenerate region, the so-called L3.5 in *C. tentans* (24), absent in *C. pallidivittatus*.

The L1-L3 subrepeat pair could have differentiated like L2-L4. L1 reminds us of the differentiated L2 in having an A-rich 3′ end (13 successive As in clone A). Obvious target sites cannot be seen but could have become obliterated by mutations.

Sequence of clone B. Clone B (Fig. 2) is similar to clone A, i.e., region L2 is largely identical to L4 and both regions are consequently parts of large subrepeats (Fig. 1d). Clone B is at the end of a tandem array, containing sequences without homology to known repeats at the 3′ end. The clone starts at the middle of subrepeat IIa, where two base substitutions have created an *Nla*III site. It then extends over L2, Ib, L3, IIb, L4, and Ia to lose identity with clone A (or the standard repeat) in the beginning of L1. The clone ends with stretches of A+T-rich, irregularly repeated DNA. One interesting feature of this clone is that it contains well-defined degenerate segments of a type that has previously been described in *C. pallidivittatus* (Fig. 1d). One such region, termed ds1, covers the distal part of IIb and extends a few base pairs into L4 (9). Another region, ds4, is situated in the central part of L4 (11). Whereas in clone A the *Nla*III cloning site is in the middle of L3, this site has been destroyed in clone B, where the sequence is identical to the *C. pallidivittatus* sequence. The B clone also has several base substitutions compared with clone A in the region preceding the flank (Fig. 2).

Clone B confirms the general structure of clone A. It also suggests that such a structure might have ancient features because of its association with flanks. Finally, it tells us that ds regions may be evolutionarily older than the time for speciation of *C. tentans* and *C. pallidivittatus*.

Formation of a 350-bp, structurally differentiated repeat from a 165-bp unit. We can now formulate a model for the evolution of the complex 350-bp repeat from a simpler 165-bp unit (Fig. 3). A unit of this size still exists in *C. thummi thummi* (8). The simple monomeric unit has a base composition that might make the DNA particularly susceptible to insertion of retroposons in agreement with a model by Rogers (27) in which T- or T+G-rich sequences could be used after a nick in the DNA to prime an mRNA from the poly(A) tail. This would result in the insertion of a sequence ending with an A-rich tail in the same strand as observed for L1 and L2. In our case, we believe that the insertion is accompanied by deletions. In a second step, a repeat was amplified together with an adjoining repeat of the original constitution. This gave rise to the type A, 350-bp repeat. Consequently, the nonsubstituted region in the previously adjoining 165-bp repeat, now without an internal homolog in the newly formed 350-bp repeat, became a linker region, i.e., future L3. A third step was the change in the ancestral L2, discussed previously, in which it lost most of its identity with L4. The outcome was a repeat of the type existing in *C. pallidivittatus*. Further differentiation could, however, take place in the form of short degenerated segments, ds regions, which represent accumulations of mutations rather than

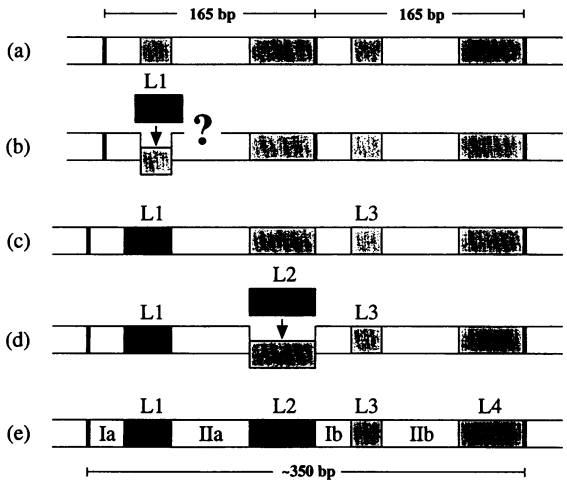

FIG. 3. Model indicating the formation of a differentiated 350-bp repeat from a simpler unit. The diagram in panel a shows the 165-bp unit, where light grey boxes indicate future linker regions between subrepeats. A hypothetical insertion of the L1 linker region and a concomitant deletion of sequence corresponding to L3 is indicated in panel b, which leads to an approximately 350-bp unit in panel c. A putative transposition of DNA into L2 with concomitant deletion of sequence with identity to L4 is indicated in panel d, leading to the differentiated 350-bp unit in panel e.

recombinations. As shown by clone B, ds regions might have already formed before the third step. In *C. tentans*, there is such a degenerate region in the middle of subrepeat IIb. This region might, however, have been generated at a later stage since a corresponding region has never been observed in *C. pallidivittatus*.

Derivation of primordial 165-bp TA unit from simple sequence DNA. According to our model, the second half of clone A probably is more representative of an ancestral 165-bp unit than the first half, and it was therefore subjected to further study. It contains two long regions with a high C/G ratio in the strand chosen for presentation (Fig. 4). These regions alternate with two stretches with a low C/G ratio, one of about 15 bp within Ib and one of about 30 bp at the end of IIb. The latter regions have more or less inverted proportions of G and C compared with the DNA with a high C/G ratio and are probably inversions, as will be shown.

Inspection of the DNA sequence in the 165-bp unit outside of the two regions with an inverted C/G ratio suggests that most of it derives from short subrepeats. Starting at the end of L3, there are five rather evenly spaced degenerate subrepeats up to the beginning of the region with a low C/G ratio,

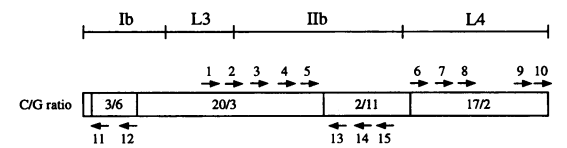

FIG. 4. Short subrepeats within the second half of clone A. On top is shown a diagram of the linker subrepeat regions. The unit is divided into regions according to C/G ratio; white boxes show regions with high ratios, and grey boxes show regions with low ratios. The positions of the different subrepeats and their orientations are given above and below the boxed diagram. The subrepeats have the numbers given in Table 1.

TABLE 1. Short subrepeats in a 165-bp primordial TA repeat

	Forward readings	Inverted readings
	CCTCAT (1)	TCTCAG (11)
	CCTAAA (2)	GCTCTA (12)
	CGTCAT (3)	CCTCAA (13)
	CCTCAA (4)	TATAAA (14)
	TCTCAA (5)	CATCGA (15)
	TCTCAT (6)	
	TCTGCA (7)	
	TTTCAC (8)	
	CCTTAA (9)	
	TCTCAA (10)	
Consensus:	(T/C) CTCAA	(T/C) CTCAA
Inverted:	TTGAG (G/A)	TTGAG (G/A)

after which follow three similar units. Then there is a short region of 14 bp in which subrepeats are lacking. At the end of L4, there are another two units. The 10 subrepeats, shown in Table 1, share the consensus (C,T)CTCAA. All these subrepeats have at least a four-of-six base agreement with the consensus. There are one to four bases between adjoining units, with an average of two. There is relatively little consensus for bases lying between the six-base units except that the units are preceded by T in five cases. Present data suggest an origin for the 165-bp unit in repeats, 7 to 10 bases long, of which 6 or 7 bases can be reconstructed. The sequence above or its complementary version, TTGAG (G,A), is similar to DNA at the ends of eukaryotic chromosomes such as *Tetrahymena* species (TTGGGG) (6), *Physarum polycephalum* (TTAGGG) (16), *Plasmodium berghei* [TT(C/T)AGGG] (25), or vertebrates (TTAGGG) (23).

Short subrepeats having at least a four-of-six base agreement with the consensus above can also be found in the two regions believed to be inverted. The first short region contains two and the second, larger region contains three short subrepeats (Table 1). The three intersubrepeat distances are between two and four bases, similar to the distances between the subrepeats in the forward direction.

Statistical significance of subrepeat pattern. An alternative explanation for the pattern of short subrepeats is that the DNA shows a bias in distribution of C and G between the two strands for some other reason and that we select short repeats from a random pattern. Closer inspection of the individual hexanucleotide repeats shows that they are of three kinds—they either agree with the consensus or differ at one or two nucleotide positions. In a DNA like the present one, containing essentially only C, A, and T in one of the strands (actual composition of the C-rich strands in the four segments of Fig. 4 is 34.5% A, 32.7% C, 26.7% T, and 6.1% G), there are two possible consensus short repeats, 21 different hexanucleotides giving one difference, and 90 different hexanucleotides showing two differences. If the short repeats represent a selection from a random sequence, we would, therefore, expect that they are present in the proportion 2:21:90 for repeats with zero, one, and two deviations, respectively. The real distribution among the 15 hexanucleotides is 4:5:6. To investigate whether this distribution is statistically different from the one expected on the basis of a random distribution, we put together the classes for zero and one deviation and compared the distribution found, 9:6, with the one expected on the random hypothesis, 3:12. Applying the binomial theorem, we added the probability for a 9:6 distribution to the probabilities for all other, less likely

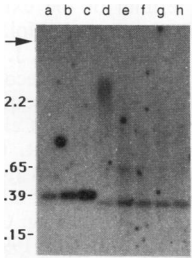

FIG. 5. Southern blot in 1% agarose gel of clone B DNA (lanes a to c) and genomic DNA from *C. pallidivittatus* (lane d) and *C. tentans* (lanes e to h). Plasmid DNA containing clone B was digested to release the insert and applied in amounts corresponding to 1,000 (lane a), 2,000 (lane b), and 4,000 (lane c) sequences per genome. *Chironomus* DNA was applied in 1-μg quantities after cleavage with *Hin*dII (lane d) and *Hin*fI (lanes e to h), both of which cleave the larger part of the TA repeat arrays to monomers (24, 29). Before cleavage, genomic DNA was treated with *Bal* 31 for 0 (lane e), 10 (lane f), 20 (lane g), or 30 (lane h) min. *Hin*fI-cleaved pBR328 DNA was used as an internal marker to measure the rate of digestion, which was 27 bp per end per min. *Bal* 31 (0.7 U/100 μl) was used in 600 mM NaCl–12 mM $CaCl_2$–12 mM $MgCl_2$–1 mM EDTA–20 mM Tris-Cl buffer (pH 8.0) at 30°C. Size markers to the left are in kilobases. Arrow shows origin.

distributions. We obtained a total probability, $P < 0.001$, that the distribution of repeats is due to selection from a random base sequence. This conclusion is reinforced by another circumstance. All repeats representing the consensus are preceded by a T. Consequently, there are three such sequences in the 118-bp sequence having a high C/G ratio. If the parts believed to be inverted are included (in the reverse orientation), there are four sequences within 165 bp. Given one such sequence, the probabilities for another two or three by chance in the 118- or 165-bp stretch of DNA are in both cases <0.01. Our data suggest that the most likely ancestral short repeat unit is T(T,C)CTCAAN_{0-3}, which has subsequently been changed by point mutations in several of the repeats.

Search for arrays of simple repeats elsewhere in the genome. Are short repeats of the present kind present in the genome also as simple sequence, truly telomeric DNA? Due to the degeneracy of the repeats, it is not possible to test by hybridization techniques all possible variants. We could, however, test for a 15-nucleotide combination which was based on the assumption that a (C,T)CTCAA motif is preceded by a T. Two such heptanucleotide sequences were joined by a degenerate position as follows: T(C,T)CTCAA (C,A,T)T(C,T)CTCAA. In clone A there are two and in clone B and the standard clone there is one 15-nucleotide sequence complementary to this degenerate oligonucleotide. *C. tentans* DNA was digested with *Hin*fI, which appears once in all kinds of repeats, electrophoresed, blotted, and hybridized to end-labelled degenerate 15-mer oligonucleotide. Hybridization occurred over the 350-bp band, at the dimer position, and as a smear coincident with the staining of the DNA. This smear is unlikely to be due to telomeric repeats since the size did not decrease after application of *Bal* 31. Furthermore, hybridization also coincided with a different staining pattern obtained after digestion of *C. pallidivittatus* DNA with *Hin*dII (Fig. 5). These results of course do not exclude that other variants of short repeats are present terminally in the chromosomes as long uninterrupted arrays.

Regions not containing short subrepeats. In the second half of clone A with noninverted proportions between C and G, two short regions do not contain obvious degenerate subrepeats. One of these regions, situated in L4 and with the sequence CCATACAAACCATC, shows an identity in its initial 9 bp with the beginning of L1. The first 7 bp of this sequence, together with 3 bp in the upstream adjoining degenerate short subrepeat, forms a 10-bp sequence, CAC CCATACA, which represents a central part of the 13-bp binding motif for the telomere binding protein in yeast RAP1 (7). This is an interesting similarity which could represent another link between the telomeric simple sequence DNA and the TA DNA in insects.

DISCUSSION

TA repeats evolved from simple sequence DNA similar to telomeric repeats. Here we reconstructed some of the evolutionary history of the 350-bp TA repeat in *Chironomus tentans* and conclude that it is highly likely to originate from short, on the average, 7- to 10-mer repeats with striking similarity to telomeric simple sequence DNA from other eukaryotes. The repeats are not evenly spaced as in several eukaryotes but vary slightly in length, more like telomeric repeats in *Saccharomyces cerevisiae* (31) or *Dictyostelium discoideum* (15). We believe that a first step led to a 165-bp unit after inversions in the arrays of short repeats. The next step might have been insertion of retroposon DNA to become L1 with deletion of preexisting DNA with identity to L3. Therefore, the size of the unit receiving the insertion has not been much changed. When, however, such a unit has become amplified together with an adjoining unit of the original 165-bp construction, a new repeating unit of 350 bp has been formed, essentially our clone A. A second transposition-deletion step may have led to further differentiation and the structure of the present-day repeat.

There is, however, also differentiation at another level. Clusters of mutations form short degenerate regions, ds regions, that characterize different subfamilies and have a characteristic intertelomeric distribution as studied in *C. pallidivittatus*. In *C. tentans*, such a cluster of mutations is localized within a subrepeat in the main repeat variant forming a short linker region.

The original asymmetry in distribution of Cs and Gs is retained in part of the repeat unit, opening the possibility that original functions may have been conserved. The further differentiation into subfamilies (9, 11) with specific intertelomeric distribution (10) could indicate the acquisition of new functions.

Distribution of TA repeats among dipteran species. HeT DNA, which is a DNA that can be added to terminally deleted chromosomes in *Drosophila melanogaster* by a unique transposition mechanism (3, 4) and is also a normal component of telomeres and pericentric heterochromatin (32, 34), shows no obvious similarities to *Chironomus* TA repeats. This is in contrast to 180-bp regular tandem repeats associated with telomeres in *Drosophila tristis* and *Drosophila obscura* (1). Also in this latter type of DNA there are traces of arrays of degenerate short repeats. It is reasonable to suggest that the *Chironomus* type of TA repeats represents an evolutionarily old form, widely distributed among dipteran species.

Clones A and B share traits of evolutionarily old TA repeats. One question is whether L2 is identical (or similar) to L4 in clones A and B not because of an ancient state but because of recombinations between standard repeats that place L4 in the position of L2. Such a construct is not possible after a

single crossover. This is because L2 is distal to L1 and proximal to L3, two regions that are distinctly different, whereas L4 comes after L3 but before L1. Consequently, a double-crossover event would be necessary. Two such independent double crossovers appear unlikely. Nevertheless, a double crossover at this level is indistinguishable from gene conversion, and one might suggest that L2 could have become converted by L4. Gene conversion can, however, only occur if there is an initial similarity, sufficient for heteroduplex formation, and is unlikely to operate between two essentially nonhomologous sequences like L2 and L4 in the standard repeat (2, 13, 20). Furthermore, the sequences in clones A and B differ from those of the standard type at several other points. We conclude that clones A and B probably represent phylogenetically old forms of repeats.

Are the TA repeats truly telomeric? We suggest that the 350-bp TA repeats have evolved from simple 7- to 10-mer repeats, six positions of which can be deduced and which show striking similarity with established eukaryotic telomeric repeats. An obvious question is whether the *Chironomus* repeats extend all the way to the end of the chromosomes. Since, however, the TA repeats are present in blocks of up to 75 kb (10), it is not easy, using, e.g., *Bal* 31 digestions, to prove such a localization. Such analyses would not tell us more than we already know from cytological observations, i.e., that these sequences are located close to the chromosome ends. Here we find that a set of short repeat variants are not present as long arrays at the chromosome ends in measurable amounts.

If we accept the assumption that the complex TA repeats reach the ends of the chromosomes, we are faced with the problem that complex DNA repeats cannot be added by telomerase. Alternatives exist, however, in unequal crossing over combined with selection for longer recombinants (22), in gene conversion (33), or, as described for *D. melanogaster*, in transposition of DNA onto chromosome ends (3, 4). All in all, we suggest that the TA repeats have evolved from simple sequence, truly telomeric DNA, retaining some of the original features and presumably functions. Evolution has added new types of sequences and probably possibilities to accommodate new functions.

ACKNOWLEDGMENTS

We are grateful to B. O. Bengtsson, T. Clementz, and B. Widegren for help with computer and statistical analysis.

This work was supported by grants from the Swedish Cancer Society and the Philip-Sörensen Foundation.

REFERENCES

1. **Bachmann, L., M. Raab, and D. Sperlich.** 1990. Evolution of a telomere associated satellite DNA sequence in the genome of *Drosophila tristis* and related species. Genetica **83**:9–16.
2. **Baltimore, D.** 1981. Gene conversion: some implications for immunoglobulin genes. Cell **24**:592–594.
3. **Biessmann, H., J. M. Mason, K. Ferry, M. d'Hulst, K. Valgeirsdottir, K. L. Traverse, and M.-L. Pardue.** 1990. Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends in *Drosophila*. Cell **61**:663–673.
4. **Biessmann, H., K. Valgeirsdottir, A. Lofsky, C. Chin, B. Ginther, R. W. Levis, and M.-L. Pardue.** 1992. HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*. Mol. Cell. Biol. **12**:3910–3918.
5. **Blackburn, E. H.** 1991. Structure and function of telomeres. Nature (London) **350**:569–573.
6. **Blackburn, E. H., and J. G. Gall.** 1978. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. J. Mol. Biol. **120**:33–53.
7. **Buchman, A. R., W. J. Kimmerly, J. Rine, and R. D. Kornberg.** 1988. Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, and telomeres in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8**:210–225.
8. **Carmona, M. J., G. Morcillo, R. Galler, E. Martinez-Salas, A. G. de la Campa, J. L. Diez, and J.-E. Edström.** 1985. Cloning and molecular characterization of a telomeric sequence from a temperature-induced Balbiani ring. Chromosoma **92**:108–115.
9. **Cohn, M., and J.-E. Edström.** 1991. Evolutionary relations between subtypes of telomere-associated repeats in *Chironomus*. J. Mol. Evol. **32**:463–468.
10. **Cohn, M., and J.-E. Edström.** 1992. Chromosome ends in *Chironomus pallidivittatus* contain different subfamilies of telomere-associated repeats. Chromosoma **101**:634–640.
11. **Cohn, M., and J.-E. Edström.** 1992. Telomere-associated repeats in *Chironomus* form discrete subfamilies generated by gene conversion. J. Mol. Evol. **35**:114–122.
12. **Cowman, A. F., R. B. Saint, R. L. Coppel, G. V. Brown, R. F. Anders, and D. J. Kemp.** 1985. Conserved sequences flank variable tandem repeats in two S-antigen genes of *Plasmodium falciparum*. Cell **40**:775–783.
13. **Dasgupta, C., and C. M. Radding.** 1982. Polar branch migration promoted by recA protein: effect of mismatched base pairs. Proc. Natl. Acad. Sci. USA **79**:762–766.
14. **Eckert, R. L., and H. Green.** 1986. Structure and evolution of the human involucrin gene. Cell **46**:583–589.
15. **Emery, H. S., and A. M. Weiner.** 1981. An irregular satellite sequence is found at the termini of the linear extrachromosomal rDNA in *Dictyostelium discoideum*. Cell **26**:411–419.
16. **Forney, J., E. R. Henderson, and E. H. Blackburn.** 1987. Identification of the telomeric sequence of the acellular slime molds *Didymium iridis* and *Physarum polycephalum*. Nucleic Acids Res. **15**:9143–9152.
17. **Greider, C. W., and E. H. Blackburn.** 1985. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. Cell **43**:405–413.
18. **Greider, C. W., and E. H. Blackburn.** 1989. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. Nature (London) **337**:331–337.
19. **Höög, C., B. Daneholt, and L. Wieslander.** 1988. Terminal repeats in long repeat arrays are likely to reflect the early evolution of Balbiani ring genes. J. Mol. Biol. **200**:655–664.
20. **Liskay, R. M., A. Letsou, and J. L. Stachelek.** 1987. Homology requirement for efficient gene conversion between duplicated chromsomal sequences in mammalian cells. Genetics **115**:161–167.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Morin, G. B., and T. R. Cech.** 1988. Mitochondrial telomeres: surprising diversity of repeated telomeric DNA sequences among six species of *Tetrahymena*. Cell **52**:367–374.
23. **Moyzis, R. K., J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff, and J. R. Wu.** 1988. A highly conserved repetitive DNA sequence, (TTAG GG)n, present at the telomeres of human chromosomes. Proc. Natl. Acad. Sci. USA **85**:6622–6626.
24. **Nielsen, L., E. R. Schmidt, and J.-E. Edström.** 1990. Subrepeats result from regional DNA sequence conservation in tandem repeats in *Chironomus* telomeres. J. Mol. Biol. **216**:577–584.
25. **Ponzi, M., T. Pace, E. Dore, and C. Frontali.** 1985. Identification of a telomeric DNA sequence in *Plasmodium berghei*. EMBO J. **4**:2991–2995.
26. **Renkawitz-Pohl, R., and S. Bialojan.** 1984. A DNA sequence of *Drosophila melanogaster* with a differential telomeric distribution. Chromosoma **89**:206–211.
27. **Rogers, J. H.** 1985. The origin and evolution of retroposons. Int. Rev. Cytol. **93**:187–279.
28. **Rubin, G. M.** 1978. Isolation of a telomeric DNA sequence from *Drosophila melanogaster*. Cold Spring Harbor Symp. Quant. Biol. **42**:1041–1046.

29. **Saiga, H., and J.-E. Edström.** 1985. Long tandem arrays of complex repeat units in *Chironomus* telomeres. EMBO J. **4**:799–804.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
31. **Shampay, J., J. W. Szostak, and E. H. Blackburn.** 1984. DNA sequences of telomeres maintained in yeast. Nature (London) **310**:154–157.
32. **Valgeirsdottir, K., K. L. Traverse, and M. L. Pardue.** 1990. HeT DNA: a family of mosaic repeated sequences specific for heterochromatin in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **87**:7998–8002.
33. **Wang, S.-S., and V. A. Zakian.** 1990. Telomere-telomere recombination provides an express pathway for telomere acquisition. Nature (London) **345**:456–458.
34. **Young, B. S., A. Pession, K. L. Traverse, C. French, and M.-L. Pardue.** 1983. Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. Cell **34**:85–94.